2. Demographic Projections

2.1. Background

The overall objective of Task 1 is to project demographic transitions for 1990, 1991, 1992, and 1993 SIPP respondents born in 1926-65, including marriage, divorce, widowhood, and mortality.[1] Figure 2.1 illustrates the states of interest and the transitions between them.

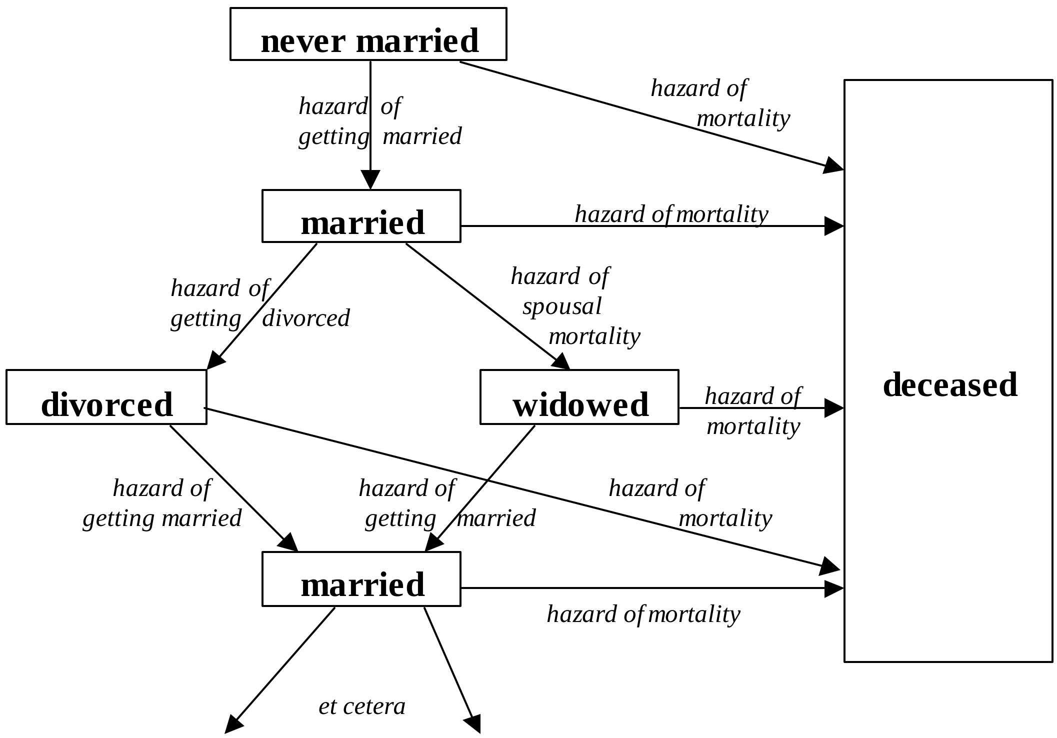

Figure 2.1. Demographic States and Transitions

As shown in Figure 2.1, there are four types of transitions:

- Marriage and remarriage
- Divorce
- Transition into widowhood (spousal mortality)
- Transition to deceased (own mortality)

In addition to marital and survival status, we project the date of onset of disability (not shown in Figure 2.1). Each transition is the outcome of a hazard process, namely the hazards of (re-)marriage, divorce, own and spousal mortality, and onset of disability.

---

[1] The Statement of Work extends to the 1990 and 1991 SIPP only and restricts the simulation sample to 1931-60 birth cohorts. The projections as described in this document and delivered are a superset of those required by the Statement of Work.

Figure 2.2 illustrates the steps that were involved in producing the demographic projections. First, we estimated model parameter coefficients of the marriage, divorce, mortality, and disability model equations. These models are based on data from 1901-1994 Vital Statistics, the 1968-1994 Panel Study of Income Dynamics, and the 1990 and 1991 waves of the Survey of Income and Program Participation (SIPP). Second, we selected the simulation sample and prepared the data. The simulation sample is based on the 1990, 1991, 1992, and 1993 SIPP waves. Third, we projected respondents' demographic transitions and future states, starting at the last survey date and ending at the time of mortality. These projections take into account the known dates of death between the last survey date and mid-1998, as recorded in SSA's Numident data.

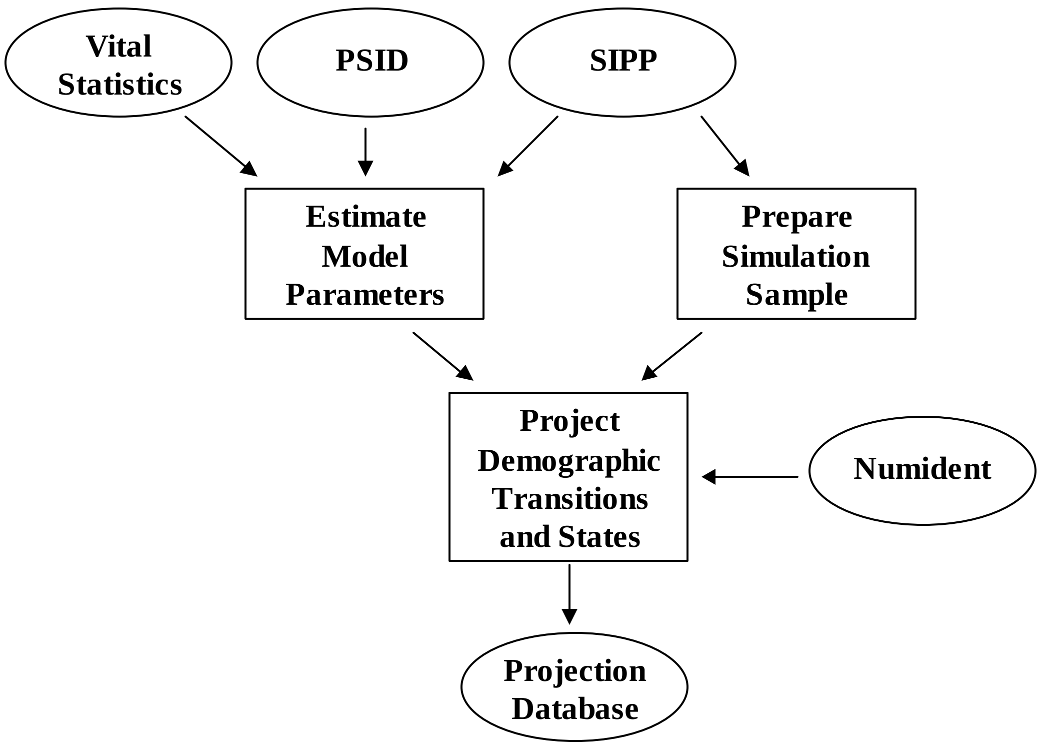

Figure 2.2. Projection Procedure Flow Chart

The sections below describe the estimation procedures and parameter estimates for own and spousal mortality (Section 2.2), marriage and remarriage (Section 2.3), divorce (Section 2.4), and the onset of disability (Section 2.5). Section 2.6 specifies the simulation sample selection criteria and discusses important data preparation issues. Section 2.7 explains the algorithms for projecting demographic states. Section2.8 presents summary statistics of the projections.[2] Appendix A documents the sequence of SAS programs that prepared the data and projected future demographic states.

[2] Chapter 4 compares aggregate projections produced by the MINT model to those produced by other demographic projection models.

2.2. The Model for Mortality

2.2.1. Overview

This section describes estimation of the mortality process parameters. Demographic projections require mortality processes for both the respondent and his or her current and future spouses.

For the current projections, mortality risk is determined by respondents' age, gender, race and ethnicity, educational attainment, permanent household income, and marital status. In addition, the projection method takes account of a secular trend towards increased longevity.

The Survey of Income and Program Participation (SIPP) offers only limited information on dates of death. A relatively small number of respondents dies during the panel period. Dates of death of 1990-93 SIPP respondents between the end of the panel and 1998 are available from administrative records in the Numident file. Mortality specifications that do not involve time-varying covariates may be estimated on these matched administrative records. However, we wish to estimate mortality as a function of time-varying marital status, on which no information is available after the end of the panel. The SIPP/Numident data therefore do not support estimation of our mortality model. Instead we estimate mortality using the Panel Study of Income Dynamics (PSID), a large household survey which has been fielded annually since 1968.[3]

While projections will only be made for respondents born in 1926-65, we estimate mortality models on all cohorts born in or before 1965. The inclusion of older cohorts is important to obtain parameter estimates for elderly persons. The 1926-65 birth cohorts need to be simulated through the year 2020, when the eldest individuals are over 90 years old.

Even though the PSID was designed to be representative of the American population, there may be differences between PSID mortality experiences and those documented in Vital Statistics of the United States. We model such differences (as a function of age, sex, race, and calendar time) and apply a procedure to transform the estimated parameters into parameters that yield projections consistent with Vital Statistics; see below. The resulting mortality hazard parameters are used to project both respondent mortality and spousal mortality (respondent transition into widowhood).

[3] We chose the PSID because it has been running for many years, has good information on deaths, marital transitions, and income, and spans the full age range.

2.2.2. The Basic Mortality Pattern

Consider Figure 2.3, which plots the natural logarithm of age-specific mortality rates (log-hazard) for white and black males and females based on 1994 Vital Statistics (National Center for Health Statistics, 1998). Mortality rates decrease sharply during the first twelve years of life, increase during adolescence, stabilize during the early twenties, and increase almost linearly from approximately age 30. The youngest members of our projection sample are around thirty as of the last survey wave, so for our purposes, the baseline log-hazard is almost linear (almost Gompertz). There is some indication in the literature that the mortality log-hazard levels off slightly at higher ages, so we allow for a piecewise linear baseline duration dependency: linear between age 30 and 65, and again linear after age 65.[4]

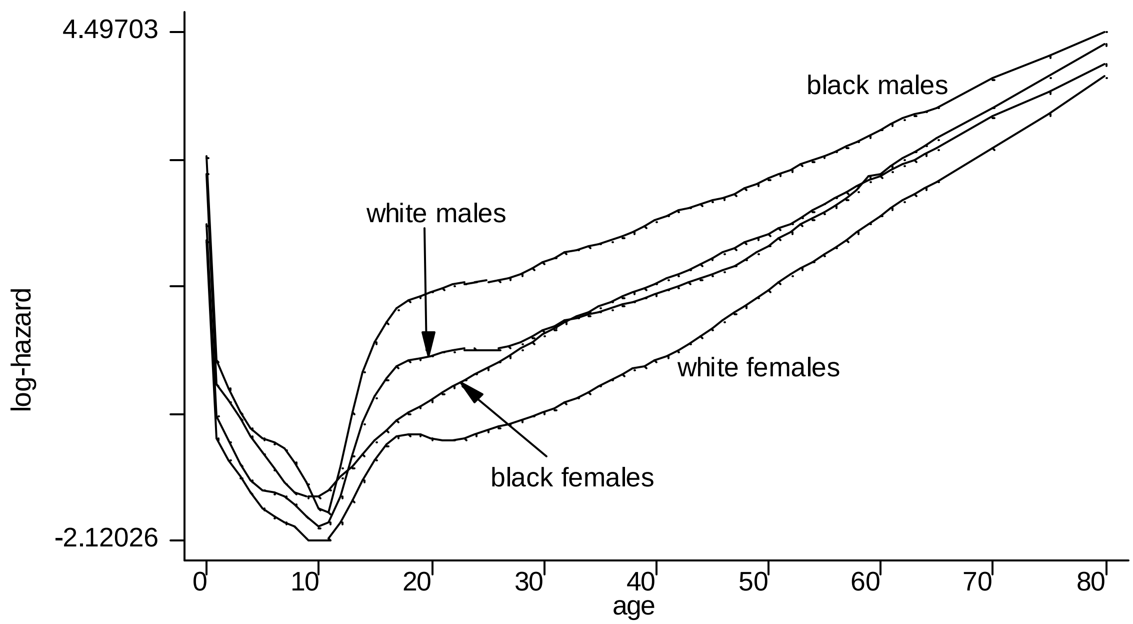

Figure 2.3. Log Death Rates, 1994 Vital Statistics

In line with the literature, we assume that the mortality process follows the standard proportional hazard model (e.g., Kalbfleisch and Prentice 1980):

$$\ln h^m(t) = g'T(t) + b'X_t, \qquad [2.1]$$

where $\ln h^m(t)$ denotes the log-hazard of dying at time t; $g'T(t)$ captures the piecewise-linear age dependency and a linear calendar time trend; and $b'X_t$

---

[4] Various studies suggest a change in the mortality function around age 90, which would call for a change in the slope of the mortality hazard line. Neither the PSID nor Vital Statistics offer sufficient richness to reliable estimate departures above age 90 from the piecewise Gompertz. As noted by Christopher Bone, because of the limited duration of MINT and the cohorts under study, this does not raise any significant issues for this implementation of MINT. The oldest individuals are only in their late eighties by the year 2020, the end of the MINT projection period. It does imply, however, that projections much beyond the year 2020 need to be interpreted with caution.

represents the effects of exogenous covariates: race, educational attainment, marital status, and permanent income.[5] The models are estimated separately for males and females. Measurement of permanent income is described in more detail below. Marital status is a time-varying covariate. Since we are only interested in projecting mortality for individuals who are at least 30 years old, we estimate the model only on PSID respondents age 30 and over. By excluding survival experiences prior to age 30, we avoid the need to carefully account for the irregular log-hazard pattern before age 30, as shown in Figure 2.3. For example, a PSID respondent who was 20 years old as of the first survey in 1968 is included in the estimation sample only starting at his 30th birthday, in 1978 (unless he died or left the sample before 1978, in which case the person does not contribute to the estimates.) Table 2.1 presents the parameter estimates.

Table 2.1. PSID Mortality Hazard Estimates

	Males	Females
Constant	-9.6619 ***	-10.0891 ***
	(.2603)	(.3314)
Age slope 30-65	.0879 ***	.0869 ***
	(.0044)	(.0057)
Age slope 65+	.0793 ***	.0867 ***
	(.0042)	(.0048)
Calendar time	-.0119 ***	-.0152 ***
	(.0038)	(.0047)
Black	.1768 **	.3219 ***
	(.0804)	(.0953)
High school drop-out	.3778 ***	.0934
	(.0704)	(.0778)
College graduate	-.0513	-.2514 *
	(.1040)	(.1427)
Never married	.2138 *	.0184
	(.1132)	(.1421)
Divorced	.4343 ***	-.1185
	(.1146)	(.1527)
Widowed	.1080	-.0041
	(.0905)	(.0805)
Permanent income	-.1591 ***	-.2675 ***
	(.0435)	(.0477)
Income missing	-.4083	-2.1304
	(1.1828)	(3.9271)
Log-Likelihood	-14424.95	

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

[5] The model of mortality does not control for disability status. The effect of disability, however, is partially captured through our control for permanent income. Also see Subsection 2.9.

All estimated patterns are consistent with well-established findings in the literature. The estimates show that net mortality rates decreased by approximately 1.19 percent (males) and 1.52 percent (females) between 1968 and 1994.[6] Blacks experience significantly higher mortality rates than whites; mortality rates decrease with educational attainment; never married and divorced men face higher mortality rates than married and widowed men, while marital status has almost no effect on women; and mortality risks are lower for individuals with higher incomes.

2.2.3. Differences Between the PSID and Vital Statistics

The PSID was designed to be representative of the American population at the time of its first wave in 1968. Since then, the immigrant composition of the US has changed and there has been some attrition from the PSID. Thus, the PSID may no longer be fully representative of the population. In addition, the PSID interview staff may not be fully successful in recording all deaths, perhaps classifying some deaths as panel attrition. For these reasons, we correct PSID mortality estimates such that they become representative of the American population and so they may be used for projection purposes.

This correction is based on a comparison of PSID mortality and mortality recorded in Vital Statistics of the United States. We collected Vital Statistics data at roughly 10-year intervals between 1901 and 1994 and converted them into mortality hazard spells, similar to the PSID data format. We then estimated mortality hazard models for individuals age 30 and over, using only sex, age, calendar time, and race as determinants. The same specification was run on PSID data. Table 2.2 presents the results. The first column shows estimates based on Vital Statistics; the second on the PSID; and the third, their difference.

Note that estimates based on Vital Statistics have very small standard errors. The reason is that they are weighted by the US population.[7]

Note that we capture mortality reductions over time by a linear trend. SSA's Office of the Chief Actuary (OACT) documents that longevity gains have varied considerably across subperiods of this century. The gains were relatively large between 1968 and 1982, and relatively small between 1982 and 1994 (Bell 1997; see Table 4.4 on page 94). One may debate whether future longevity gains will follow the pace of the entire period since the beginning of this century, or since the establishment of Medicare in the late 1960s, or even since more recent dates. Only time will tell. We take a very long term view and extrapolate from the beginning of this century.

---

[6] The trend in Vital Statistics mortality rates, which will be used for projection purposes, is slightly flatter. See below.

[7] The weights have been divided by 1000, so that standard errors are in fact 1/1000-th of those presented here.

Table 2.2. Differences Between PSID and Vital Statistics

	VS	PSID	PSID-VS
Males			
Constant	-8.3597 ***	-9.6791 ***	-1.3195 ***
	(0.0013)	(0.2480)	(0.2480)
Age slope 30-65	0.0721 ***	0.0909 ***	0.0187 ***
	(0.0000)	(0.0042)	(0.0042)
Age slope 65+	0.0821 ***	0.0838 ***	0.0017
	(0.0000)	(0.0039)	(0.0039)
Time 1901-1994	-0.0081 ***	-0.0179 ***	-0.0099 ***
	(0.0000)	(0.0037)	(0.0037)
Black	0.2815 ***	0.3913 ***	0.1097
	(0.0004)	(0.0778)	(0.0778)
Females			
Constant	-8.7528 ***	-10.2761 ***	-1.5233 ***
	(0.0016)	(0.3260)	(0.3260)
Age slope 30-65	0.0685 ***	0.0902 ***	0.0217 ***
	(0.0000)	(0.0055)	(0.0055)
Age slope 65+	0.0954 ***	0.0862 ***	-0.0093 **
	(0.0000)	(0.0043)	(0.0043)
Time 1901-1994	-0.0141 ***	-0.0181 ***	-0.0040
	(0.0000)	(0.0044)	(0.0044)
Black	0.3325 ***	0.5323 ***	0.1998 **
	(0.0005)	(0.0912)	(0.0912)
Log-Likelihood	-222314824.9	-14498.74	-14498.74

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

To ensure that our mortality projections, in the aggregate, match those which would be produced by Vital Statistics estimates, we correct the PSID mortality estimates of Table 2.1 by the difference of PSID and Vital Statistics estimates, as in the third column of Table 2.2. *The mortality specification that we use to project dates of death for the SIPP sample is given by Table 2.1 minus the coefficients of the third column of Table 2.2*.

2.2.4. Measurement of Permanent Income

Our measure of permanent income is based on individuals' long-run position in the distribution of household log-income. The SIPP panels contain 32 monthly household income values (eight waves with four monthly values each). SAS program `income.sas` groups these into the first 10 values, the next 12 values, and the last 10 values. Program `perminc.sas` rescales these sums such that they represent annual values and estimates a very simple model in which annual log-income is regressed on age (piecewise linear with different slopes before and after age 50), sex interacted with marital status (never married, divorced, and widowed relative to married), and a measure of number of adult-equivalents in the household.[8] Table 2.3 shows the results of this regression.

For each respondent and each of his or her three annual incomes, we computed the residual. For each respondent, we computed the average of his or her three residuals and took this average as a measure of permanent income. The same procedure was applied to the PSID. While many more than three annual household income measures are available in the PSID, we restricted ourselves to the first three incomes after the respondent reached age 30, so as to be compatible with the SIPP measurement.

Table 2.3. Household Log-Income Parameter Estimates

Constant	9.3733 *** (0.0215)
Age slope 25-50	0.0110 *** (0.0005)
Age slope 50+	-0.0156 *** (0.0004)
Never married male	-0.1267 *** (0.0129)
Never married female	-0.3486 *** (0.0138)
Divorced male	-0.1916 *** (0.0162)
Divorced female	-0.4963 *** (0.0136)
Widowed male	-0.2016 *** (0.0267)
Widowed female	-0.3876 *** (0.0136)
log(adults equivalent)	0.7541 *** (0.0118)

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

As shown in Table 2.1, our measure of permanent income is strongly predictive of mortality risk.[9]

[8] This measure, log(adults+0.7*kids)$^{0.65}$, is based on recent research on poverty measurement, which suggests that (adults+0.7*kids)$^{0.65}$ is a reasonable conversion of adults and children in a household into adult need equivalents.

[9] An alternative measure of permanent income is the respondent' s Average Indexed Monthly Earnings (AIME), or an equivalent summary measure computed for younger workers. This measure is available in the SIPP from matched SSA records and may be computed in the PSID from self-reported information. However, years in non-covered employment cannot be distinguished from years with zero earnings in matched SSA records. This is a potentially serious limitation, especially for earlier years when Social Security coverage was far from universal.

Table 2.4 shows remaining life expectancies for a 60-year-old in 1990 by sex, race, and a combination of permanent income and education. This table is generated from parameter estimates of Table 2.1 (corrected by Table 2.2) for stereotypical values of the covariates.[10] The income points correspond to the first quartile, median, and third quartile. Our model controls for both income and education, which are highly correlated. Projections of life expectancies by income, holding education constant, would therefore understate differences by income. We therefore show projections by income, assuming that the lower incomes have less than a high school education, the median are high school graduates, and the third quartile corresponds to college graduates. "Q1 income—high school drop-out" represents a high school drop-out whose permanent income measure is equal to the first quartile cut-off; "Median income—high school graduate" represents a high school graduate with median permanent income; and "Q3 income—college graduate" represents a college graduate with permanent income equal to the third quartile cut-off.

Table 2.4. Remaining Life Expectancies at Age 60 by Sex, Race, and Income/Education

	Male	Female
White		
Q1 income—high school drop-out	16.4	23.9
Median income—high school graduate	20.6	26.1
Q3 income—college graduate	21.8	29.8
Black		
Q1 income—high school drop-out	15.7	22.7
Median income—high school graduate	19.7	25.0
Q3 income—college graduate	20.9	28.6

Note that life expectancy differences between the first and third income quartile cut-offs are between five and six years. This has important implications for poverty in old age. As projected by The Urban Institute/Brookings Institution, individuals with low lifetime income may enter retirement with limited financial resources. As projected by RAND, these resources will need to support a shorter retirement period, on average, than experienced by higher-income and better-educated individuals. It also has important implications for the degree of progressivity that is implicit in the Social Security program (Panis and Lillard, 1996).

[10] The table contains "cohort" life expectancies and may not be directly compared to standard "current" life expectancies as published in Vital Statistics publications; see Section 4.4.1 for the definition.

2.3. The Model for Marriage and Remarriage

In line with the literature, we model the transitions into marriage using a continuous time hazard model, also known as a failure-time model (e.g., Kalbfleisch and Prentice, 1980). Its basic form is given by piecewise-linear Gompertz. The multiplicative effects on covariates on the hazard are equivalent to additive effects on the log-hazard:

$$\ln h_{ij}^{w}(t) = \Gamma_w(t) + q_w' X_{ij} \qquad [2.2]$$

where $\ln h_{ij}^{w}(t)$ is the log-hazard that individual i marries (w for wedding) for the j-th time. The marriage baseline hazard, $\Gamma_w(t)$, captures duration dependencies on respondent age and duration since the previous marriage dissolved. In addition, as discussed below, $\Gamma_w(t)$ may include a duration dependency on calendar time to capture secular changes in marriage rates. All covariates are constant within spells; some, such as the number of previous marriages, differ across marriages, but do not vary over time within a spell. Throughout we suppress the person subscript.

The transition into (re-)marriage involves a period during which the individual is unmarried and "at risk" of marrying. Once married, the individual is no longer at risk of marrying. (We assume monogamy.) Instead, he/she enters a new period in which he/she is at risk of divorcing. Alternatively, the marriage may end through the death of the person's spouse. After the divorce or widowhood, the individual enters a new period in which he/she is at risk of re-marrying. The marriage and remarriage processes are thus naturally captured by hazard models, also known as failure-time models.

We do not account for unobserved heterogeneity, even though it has been shown to be significant in our own earlier work and not independent of mortality risk (Lillard and Panis, 1998b). The reasons for this exclusion here are that the projection exercise would be very much more complicated (and thus impossible to complete within the required time frame) and that it would rely on distributional assumptions that would undoubtedly be controversial. To our knowledge, no one has worked out the technique for projections of hazard processes that are based on random effects heterogeneity. For purposes of the Near Term Model, exclusion of heterogeneity is not a severe limitation. The main purpose of the Model is to yield accurate predictions, not to estimate structural parameters with behavioral interpretations. A model without heterogeneity but with extensive controls for parity (marriage number) will generate accurate predictions. We experimented extensively with parity controls, both in additive and interactive form.

2.3.1. The Data

The model may be estimated on any data set that contains longitudinal information on marriage and divorce. The SIPP itself is an excellent candidate, as is the PSID with which we have ample experience (Lillard 1993; Lillard and Panis 1996, 1998a, 1998b; Panis and Lillard 1996).[11] Since the SIPP population is the population on which projections will be based, we propose to estimate models of marriage on the SIPP panels. Only the 1990 and 1991 SIPP panels are used for estimating the (re-) marriage process; the 1992 and 1993 panels are used to assess out-of-sample goodness-of-fit; see Subsection 2.3.2 (page 27).

Marital History Data Quality Issues

Although SIPP data files are among the cleanest of all major longitudinal surveys, some data quality issues inevitably arise. We highlight the most important marriage history issues.

SIPP marriage history information is only obtained for the first two and the most recent marriage. If respondents were married more than three times, we do not know how many times exactly, or the dates when they married, divorced, and/or widowed. We imputed the number of marriages and transition types/dates based on the PSID, which contains full information. We estimated a simple ordered probit model of number of marriages, using the period between the dissolution of the second marriage and the most recent wedding date as sole explanatory variable. (No other variable was found to be predictive.) We then stochastically imputed the number of SIPP marriages based on the same gap measure. Dissolution types (divorce versus widowhood) were randomly assigned based on the fractions found to divorce (85.1 percent) and end in widowhood (14.9 percent) in the PSID. Transition dates were selected such that marriages were spread evenly between the dissolution of the second marriage and the wedding of the most recent marriage.[12]

Marriage transition dates are reported to the month only. Very short marriages and very short divorce/widowhood spells were therefore sometimes reported to result in multiple transition dates in the same month. Instead of selecting the 15-th of the month as our best-guess transition date, we chose the 10-th and the 20-th for the two dates.

We updated marriage histories as reported in the Wave 2 Topical Modules with panel information through the end of the survey sequence. In quite a few cases, the status reported for month 9 was not the same as in the Topical module. In many cases, a legitimate transition was the most likely cause. The remaining cases followed the basic rule that the marital status as of the last marriage described by the topical module was correct. The monthly series was adjusted accordingly starting in month 9 to be consistent with the last observed marital status on the topical module. Processing forward from month 9 to 32/36/40 (depending on the number of SIPP waves), we recorded any changes in marital status. The details of this consistency

[11] Our prior work focused on the timing of marital separation rather than divorce.
[12] For estimation purposes, windows were created around best-guess transition dates that were as wide as possible, so that the additional marriages contribute through their parity but very little through their timing.

adjustment are extensively documented in the source code itself (`updatemar.sas`). There were cases that transitioned from never married to divorced or widowed or from separated to married; in each case, a general rule was formulated to resolve the issue as well as possible.

In a handful of cases, respondents reported a first marriage date before their birth date. In a few dozen cases, first marriages presumably took place before age 12. We accepted such respondents' reports in the sense that we took them as baseline for the projections, but we did not use them to estimate models of getting (re-)married and divorced.

All inconsistencies were flagged by assigning non-zero values to variable `marqual`. Only "clean" marriage histories were used in estimating hazard models of getting (re-)married and divorced.

Explanatory Covariates

As is well known from the literature (including our own contributions), age, sex, education, and race/ethnicity are powerful predictors of marital status changes. In addition, the timing of a remarriage is determined by the duration since the previous marriage ended and the current marital status (divorced or widowed); the timing of a divorce is determined by the duration since the wedding. All these factors will be incorporated in the duration dependencies $\Gamma_w(t)$ and the covariates X_{ij}.

We did not control for spousal compatibility measures such as the difference in age between husband and wife, differences in race/ethnicity, and the difference in educational attainment. Spousal characteristics are not available for marriages that were completed prior to the first SIPP interview and may thus not be used for estimation purposes.

Another powerful predictor of marital transitions is the number of children that the couple has (and the number born outside marriage or brought in from prior marriages). The main problem with such measures is that their values are unknown for the projection period. One would need to develop additional models for fertility (separately for marital and nonmarital), and project future births. The issue is further complicated by evidence that childbearing is endogenous to divorce risk (Lillard 1993), so that systems of simultaneous hazard equations with correlated heterogeneity would need to be developed, estimated, and projected. This would be a huge undertaking, well beyond the scope of the current project and with only very small benefits to the current project.[13]

[13] As noted by reviewer John Rust, the policy applicability of MINT would be greatly enhanced if MINT were expanded to a closed overlapping generations model of the full U.S. population. To that end, the value of including a fertility module would be very high.

There has been a marked trend in marriage rates in the United States. Figure 2.4 shows the number of marriages per 1,000 unmarried women age 15 and over for 1940 through 1990 (NCHS 1995a) and indicates a steady decline in the marriage rate from 1947 to the current time. We control for a linear time trend in our marriage model specification to capture changes over time not accounted for by other covariates in the model.

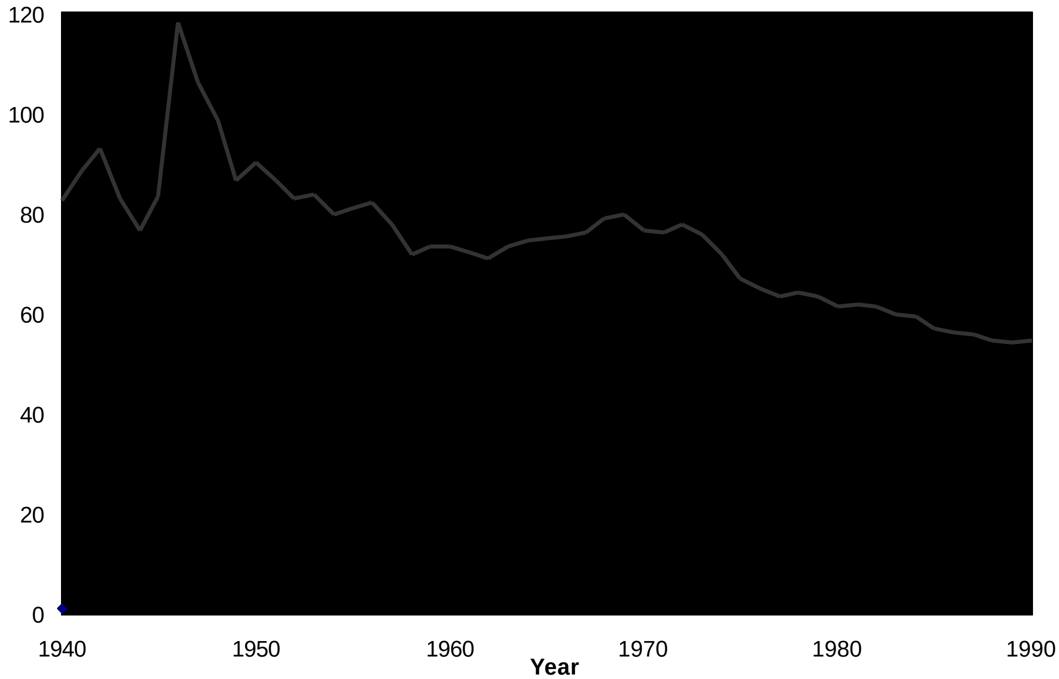

Figure 2.4. Marriages per 1,000 Unmarried Women Aged 15+, 1940-1990

Table 2.5 shows parameter estimates of the marriage and remarriage process, estimated separately for males and females. The age pattern indicates that marriage rates increase until age 20 and decrease thereafter. For remarriage, the hazard increases during the first three years after dissolution of the previous marriage, and decrease thereafter. Marriage rates are decreasing over time, consistent with Figure 2.4. The hazard of marriage for the second, third, and subsequent times are higher than for the first time. (This may be due to heterogeneity rather than marriage number; see above.) White non-Hispanic persons are more likely to enter a marriage than other races and ethnicities. Men who are high school drop-outs tend to marry later than high school graduates, whereas the pattern is reversed for women. College graduates tend to marry later than high school graduates. Men with a high permanent income, measured as explained in subsection 2.2.4 (page 19), tend to marry sooner; their female counterparts later.

Table 2.5. Estimates of Marriage Formation

	Males	Females
Constant	-23.7332 ***	-21.9557 ***
	(1.2834)	(.5813)
Age slope 0-16	1.1847 ***	1.1783 ***
	(.0813)	(.0370)
Age slope 16-20	.6211 ***	.3855 ***
	(.0121)	(.0072)
Age slope 20-25	.0840 ***	-.0545 ***
	(.0041)	(.0038)
Age slope 25+	-.0496 ***	-.0751 ***
	(.0010)	(.0012)
Slope on duration unmarried,	.1208 ***	.0789 ***
0-3 years	(.0153)	(.0146)
Slope on duration unmarried,	-.1086 ***	-.0726 ***
3-8 years	(.0101)	(.0094)
Slope on duration unmarried,	-.0382 ***	-.0223 ***
8+ years	(.0074)	(.0061)
Calendar time	-.0079 ***	-.0036 ***
	(.0004)	(.0003)
Married once before	.4325 ***	.3590 ***
	(.0327)	(.0304)
Married twice before	.6669 ***	.6248 ***
	(.0425)	(.0395)
Married three or more times before	1.2981 ***	1.2017 ***
	(.0576)	(.0506)
Black	-.3587 ***	-.5179 ***
	(.0208)	(.0183)
American Indian, Eskimo or Aleut	-.1756 **	-.0543
	(.0750)	(.0647)
Asian or Pacific Islander	-.2368 ***	-.2276 ***
	(.0491)	(.0425)
Hispanic	-.0592 **	-.3009 ***
	(.0241)	(.0232)
High school drop-out	-.0744 ***	.1284 ***
	(.0153)	(.0134)
College graduate	-.1733 ***	-.4313 ***
	(.0153)	(.0173)
Widowed	.2856 ***	-.3813 ***
	(.0399)	(.0356)
Permanent income	.0164 ***	-.0279 ***
	(.0059)	(.0049)
Log-Likelihood	-328,842.85	

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

2.3.2. Goodness of Fit of Marriage Transition Models

Our hazard models of getting married and divorced (Section 2.4, below) are based on experiences of the 1990 and 1991 SIPP respondents. We applied these estimates to 1992 and 1993 SIPP respondents to assess the goodness of fit. Starting all respondents at age 12 (when no one is married yet), we projected marital transitions until the last interview date. Table 2.6 shows actual marital status and projected marital status for these 1992 and 1993 respondents.

Table 2.6. Actual and Projected Marital Status

	Actual status		Projected status	
	Frequency	Percent	Frequency	Percent
Never married	4301	11.3	3954	10.4
Married	28065	73.7	27198	71.4
Widowed	1382	3.6	1863	4.9
Divorced	4346	11.4	5079	13.3

As is clear from the table, projected and actual marital status distributions are very close. The discrepancies may be due to stochasticity (because of duration draws in the projection method) or to a mild self-selection. The projection namely assumes that all respondents survive through the last survey. In reality, SIPP respondents are the survivors of their birth cohorts, and thus somewhat self-selected.

The distributions of projected number of marriages and age at first marriage are also very close to the actual distributions (not shown here; see `checkmar.sas`).

2.4. The Model for Divorce

Similar to the model for marriage formation, we model marriage dissolutions using a continuous time hazard model:

$$\ln h_{ij}^{d}(t) = \Gamma_d(t) + q_d' X_{ij} \quad [2.3]$$

where $\ln h_{ij}^{d}(t)$ is the log-hazard of divorcing (d) for the j-th time. The baseline hazard, $\Gamma_d(t)$, captures duration dependencies on the duration since the wedding and respondent age. In addition, as discussed below, $\Gamma_d(t)$ includes a duration dependency on calendar time to capture secular changes in divorce rates. All covariates are constant over the duration of the divorce spell; some, such as the number of previous marriages, differ across marriages, but do not vary over time within a marriage. For reasons discussed above, we do not account for unobserved heterogeneity. Throughout, we suppress the person subscript.

Marriages that end in widowhood will never result in a divorce. These marriages thus contribute censored dissolution spells. Similarly, marriages that are still in progress at the last interview date contribute a censored spell. Hazard models offer a natural way to incorporate such censored durations.

As with the model for marriage and remarriage, we use 1990 and 1991 SIPP data to estimate our model of divorce behavior. Data from the 1992 and 1993 panels were used to assess goodness-of-fit; see above.

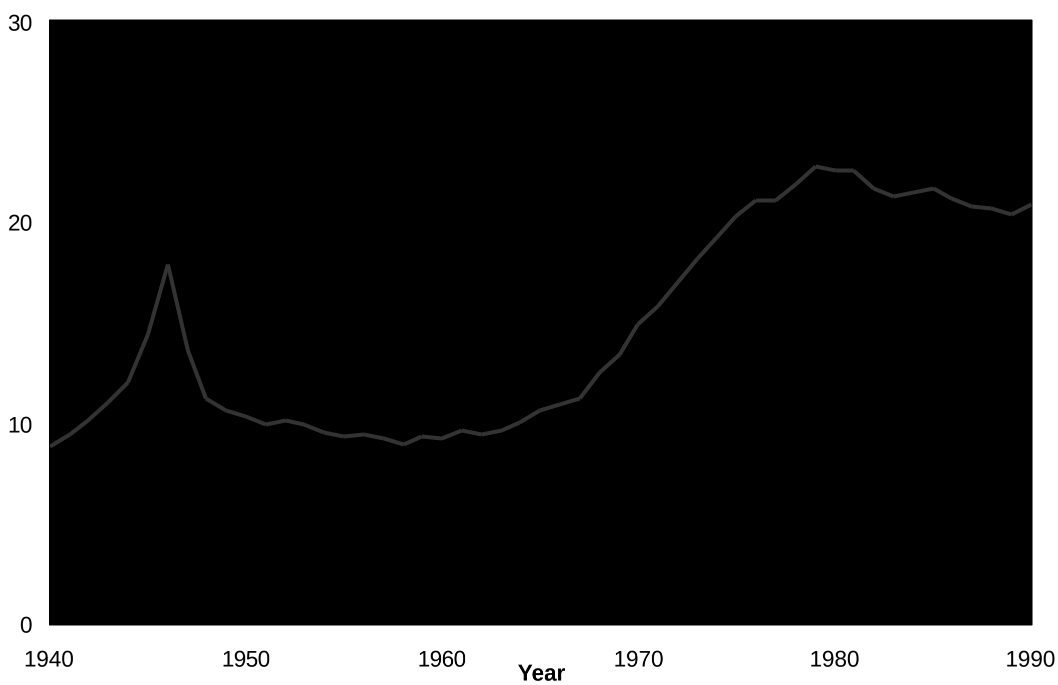

Figure 2.5. Divorce Rate per 1,000 Married Women Aged 15+, 1940-1990

Figure 2.5 shows the divorce rate per 1,000 married women aged 15 and over in the United States from 1940 to 1990 (NCHS 1995b). The divorce rate increased steadily between 1960 and 1980; since 1980, the trend has been approximately flat. We therefore include a piecewise-linear time trend in our divorce specification with a node at 1980. Table 2.7 shows the parameter estimates.

Table 2.7. Divorce Hazard Estimates

	Male	Female
Constant	-1.0198***	-1.7268***
	(.1100)	(.0946)
Age slope, 0-30 years	-.1193***	-.1021***
	(.0038)	(.0032)
Age slope, 30+ years	-.0400***	-.0523***
	(.0015)	(.0015)
Marriage duration, 0-1 years	.4339***	.7350***
	(.0724)	(.0694)
Marriage duration, 1-4 years	.2395***	.1526***
	(.0117)	(.0107)
Marriage duration, 4-15 years	-.0228***	-.0156***
	(.0032)	(.0030)
Marriage duration, 15-25 years	-.0386***	-.0275***
	(.0048)	(.0044)
Marriage duration, 25+ years	-.0875***	-.0832***
	(.0060)	(.0052)
Calendar time, pre-1980	.0401***	.0429***
	(.0010)	(.0008)
Calendar time, post-1980	-.0025	.0058***
	(.0020)	(.0019)
Second marriage	.5737***	.6368***
	(.0248)	(.0232)
Third or higher marriage	1.2503***	1.3584***
	(.0396)	(.0338)
High school drop-out	-.0274	-.0085
	(.0208)	(.0186)
College graduate	-.2117***	-.1068***
	(.0204)	(.0215)
Black	.1198***	.1786***
	(.0276)	(.0240)
American Indian, Eskimo or Aleut	.3339***	.3237***
	(.0766)	(.0611)
Asian or Pacific Islander	-.6198***	-.6378***
	(.0692)	(.0610)
Hispanic	-.3015***	-.2076***
	(.0343)	(.0314)
Log-Likelihood	-687,975.70	

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

Table 2.7 indicates that divorce rates decrease with age. They increase during the first four years of marriage and decline as the marriage lasts longer. The estimate of the time trend parameters confirms the trend in Figure 2.5: divorce rates increased significantly until 1980 and remained almost unchanged since then. (Our projection algorithms assume that the post-1980 trend continues to the year 2020.) Divorce rates are higher for second and subsequent marriages than for first marriages. (This may be due to heterogeneity rather than marriage number; see above.) Blacks and native Americans experience higher divorce rates than whites, Asians, and Pacific Islanders. Hispanics experience lower divorce rates than non-Hispanics.

Given that husbands and wives always get divorced at the same time, we would ideally want to estimate the divorce equation at the couple level, i.e., controlling for both spouses' characteristics including spousal compatibility measures. However, spousal characteristics are only known for marriages that were ongoing during the SIPP panel. The characteristics of former spouses are unknown. It is therefore impossible to estimate the divorce equation at the couple level.

2.5. The Model for Onset of Disability

Our demographic projections do not involve health or disability status. However, The Urban Institute and Brookings Institution found self-reported functional disability to be strongly predictive of earnings. In addition to marital and survival status, we therefore project disability status.

Disability is defined as self-reported functional disability: " Does ... have a physical, mental, or other health condition which limits the kind or amount of work ... can do?" We simplify reality by assuming that disability is an absorbing state, i.e., one cannot recover. We model the timing of onset of first disability report.

The SIPP, in its Work Disability History Topical Module, asks for functional disability. If functional disability is present, the date of onset is asked. The Work Disability History Topical Module is only administered to respondents age 16-67. Our model for the onset of disability is based on pooled observations from the 1990 and 1991 SIPP. Since the main objective is to project dates of disability onset for individuals at least around 30 years of age, we only include respondents in the estimation data set that are not disabled as of their 30th birthday. In other words, the disability spells upon which our estimates are based all begin at age 30 and continue through either the date of disability onset or the interview date. Respondents that indicated being disabled but who did not provide a date of onset were excluded from the estimation sample.

Table 2.8 presents the results of estimation. The risk of becoming disabled increases with age and accelerates after one' s 45th birthday. There is no significant difference between males and females. High school drop-outs are far more likely to become disabled than high school graduates; college graduate experience even lower disability rates. Asians and Pacific Islanders face the lowest disability risks, followed by whites and blacks. Native Americans experience the highest disability rates. Individuals of Hispanic origin are less likely to become disabled than non-Hispanics.

Table 2.8. Estimates of Onset of Disability

Constant	-7.3766 ***
	(0.1786)
Age slope, 30-45	0.0526 ***
	(0.0045)
Age slope, 45+	0.1746 ***
	(0.0047)
Male	0.0062
	(0.0348)
High school drop-out	0.7312 ***
	(0.0389)
College graduate	-0.6668 ***
	(0.0577)
Black	0.2779 ***
	(0.0487)
American Indian, Eskimo or Aleut	0.5446 ***
	(0.1465)
Asian or Pacific Islander	-0.5249 ***
	(0.1378)
Hispanic	-0.1674 **
	(0.0681)
Log-Likelihood	-25736.61

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%

2.6. Data Preparation Issues

Most of the data preparation was applied to all respondents to the 1990, 1991, 1992, and 1993 SIPP panels, regardless of birth year, so that models of (re-)marriage, divorce, and disability could be estimated on all age ranges. However, demographic transitions are projected only for respondents born in 1926-65 (boundaries inclusive). The sample selection criteria are:

1. Year of birth not before 1926 and not after 1965; AND
2. Strictly positive value for full-panel person weight (`pnlwgt`) OR be present until the last interview wave.

Extensive exploration of the data revealed some puzzling issues related to person weight. First, variable `pnlwgt` is zero for approximately 15 percent of individuals present in all interview waves. Second, `pnlwgt` is often nonzero for individuals who left the sample before the full panel was administered. Third, `pnlwgt` is very frequently nonzero for 1992 panel respondents who only participated in nine of the ten 1992 interviews.

Consultation with SIPP experts Denton Vaughan and Judy Eargle indicated the following.[14] Nonzero `pnlwgt` values for individuals who did not respond to all interviews may be legitimate where the individual was deceased, entered an institution, moved into military barracks, moved abroad, or otherwise became ineligible for follow-up. The frequent occurrence of nonzero `pnlwgt` for 1992 respondents who participated in all but the last interview is explained by the government shutdown of December 1994 which forced the Census Bureau to cancel follow-up interviews with at least one rotation group. Zero `pnlwgt` values for about 15 percent of individuals who participated in all interviews remain a mystery. Regardless of the exact explanations, *the Census Bureau recommends that policy analysis should be based on cases with strictly positive `pnlwgt` only.*

Table 2.9. Simulation Sample Sizes

	`pnlwgt`		
	0	>0	Total
Birth year 1926-30	1,952	6,656	8,608
Birth year 1931-60	24,960	59,537	84,497
Birth year 1961-65	7,998	11,968	19,966
Total	34,910	78,161	113,071

Table 2.9 shows the number of observations in the simulation sample.[15] The total sample size is 113,071. Of these, 34,910 have a zero value of `pnlwgt` even though

[14] E-mail communication from Denton Vaughan to Howard Iams of January 5, 1999.

[15] The 1993 SIPP panel contains two male respondents that reported being married: IDs 7451101.11.101 and 7451101.11.102. Given the need to project the same potential divorce date for

they participated until the last interview. *The sample for analysis purposes, with strictly positive `pnlwgt` values, consists of 78,161 individuals.* Of these, 59,537 are born in 1931-60 (the cohorts that were specified in the Scope of Work); an additional 18,624 are born during the five years on either side of the 1931-60 birth cohorts.

Marital Status Issues

As discussed above, there were some issues with the quality of marital status reports. We updated marriage histories as reported in the Wave 2 Topical Modules with panel information through the end of the survey sequence. In quite a few cases, the status reported for month 9 was not the same as in the Topical Module; in other cases, respondents went through "illegal" transitions (never married to divorced, separated to married, etc.) from one month to the next. In a handful of cases, respondents reported a first marriage date before their birth date. In a few dozen cases, first marriages presumably took place before age 12.

For purposes of estimating models of marriage and divorce, we dropped individuals with poor reporting quality. For purposes of projecting future transitions and states, however, we included all respondents regardless of the quality of their reports. As a general rule, we assume that the most recent marital status report is correct, so that the projections (which start at the last interview date) are based on the most recent marital status report. The projection data set (`mint.sd2`) contains both historical and future marital transitions; the historical transitions reflect our best judgment of actual transitions.

Disability Status Issues

The SIPP records disability status and date of onset in the Work Disability History Topical Module. This module is only administered to respondents age 16-67. All MINT simulation respondents are in that age range and should have been administered the Topical Module. However, disability status is missing for 16,921 respondents in the projection sample. In addition, disability status is unknown for former spouses, i.e., individuals to whom the respondent was married, but either deceased or divorced before the SIPP panel. Furthermore, 1,697 individuals reported being disabled but did not provide a date of onset. We imputed disability status and/or onset date for these groups.

The imputation algorithms are identical to those being used for future projections of disability status and other hazard outcomes; see Section 2.7 below. For former spouses and for respondents with missing disability status, we imputed an onset date. If that date fell before the interview date, we coded variable `disabled`=1 (to indicate

both spouses, a special algorithm would need to be developed to ensure spousal consistency for this couple. Instead, we dropped this couple from the simulation data.

that he or she became disable before death) and `disabdte` to indicate the date of onset, as imputed. If the imputed date was after the last interview, we coded `disabled`=0 and left the onset date, `disabdte`, equal to missing. For projection purposes, these individuals were treated identically to respondents who indicated that they were not disabled. For the 1,697 individuals that reported being disabled, but did not provide an onset date, we imputed an onset date under the restriction that the date fall before the interview date.

Spousal Characteristics

Projections of widowhood and divorce dates require information about both own and spousal characteristics: spousal sex, date of birth, race, ethnicity, and education. In addition, for The Urban Institute/Brookings Institution to project future earnings, disability status and the date of disability onset are required. These characteristics are only known for spouses who were themselves respondents to the SIPP surveys. Even if they participated in only one interview, we recorded their characteristics.

By request of The Urban Institute/Brookings Institution, we imputed spousal characteristics for former spouses. The imputation algorithms are based on empirical couple distributions in the SIPP data. Consider imputations of race. We cross-tabulated the races of husbands and wives in the data. To assign the race of a former spouse of, say, an Asian person, we drew a uniformly distributed random number and assigned a race according to the empirical distribution of spousal races: spouses of Asian persons were white, black, Asian, and Native American in 52, 0, 46, and 2 percent of the cases, respectively. The race of a former spouse of a white, black, or Native American person was imputed in a similar manner. Similarly, a Hispanic person marries another Hispanic person in 87 percent of the cases; a non-Hispanic person marries another non-Hispanic person about 99 percent of the cases. Similarly, the education of a high school graduate's former spouse was assigned based on the finding that high school graduates marry high school drop-outs in 12 percent of the cases, high school graduates in 72 percent of the cases, and college graduates in 16 percent of the cases. Spousal dates of birth were imputed using the empirical distribution of the age difference between husbands and wives. Imputations of spousal disability status and date of onset of disability were based on the disability model, as discussed above, and not on the empirical joint distribution of husbands' and wives' disability statuses.

Spousal characteristics are recorded in array variables. For example, educational attainment of spouses are recorded in variables `speduc1` through `speduc8`, allowing for up to eight spouses (marriages) per respondent.

2.7. Projection Algorithms

Figure 2.1 (page 13) shows potential demographic transitions and the hazard processes that drive their timing. The projection method is as follows. As of the last interview wave, an individual finds himself in any one of the demographic states shown in Figure 2.1. Depending on the state, he is subject to two or more transition hazards. For example, suppose the person is never married. He may become (1) married or (2) deceased. His next state is determined by whichever transition takes place first. To this end, we generate two durations, namely until marriage and death. The various demographic states affect each others' transition hazards, but conditional on observables, all hazard processes are statistically uncorrelated, and we may generate durations independently.

The probability that a generic event has not happened yet as of time t is by definition given by its survivor function, *S(t)*, i.e., by one minus the cumulative probability function, 1-*F(t)*. The hazard is by definition the relative decline of the survivor function,

$$h(t) = -\frac{dS(t)/dt}{S(t)}, \qquad [2.4]$$

so that, still by definition,

$$S(t) = \exp\left\{-\int_{t=t_0}^{t} h(t)\,dt\right\}, \qquad [2.5]$$

where t_0 is the time at which the event became at risk of occurrence. The median duration t^m until an event occurs is given by the solution to $S(t^m) = \frac{1}{2}$, i.e., $t^m = S^{-1}(\frac{1}{2})$. Note that all hazard models in our projection exercise are of the general form

$$\ln h(t) = gT(t) + b'X_t, \qquad [2.6]$$

i.e., the log-hazard is piecewise-linear in durations *t*. This implies that there is a closed-form solution to the survivor function and also to its inverse, i.e., the expected duration may be found by a closed form computation. In addition to being very flexible, piecewise-linear duration dependencies have the advantage that all computations have a closed-form solution, i.e., no numerical integration is required.

For purposes of projecting dates of death and other demographic transitions, the expected duration is not the desired concept, since it would lead to predictions that all respondents die exactly after their remaining life expectancy. (Also see Chapter 3 on Stochastic Elements.) Instead, we draw randomly from the distribution of durations. This is accomplished by drawing a random number between 0 and 1, say, S^*, and solving for the duration t^* as

$$S^* = S(t^*) \quad \Leftrightarrow \quad t^* = S^{-1}(S^*). \qquad [2.7]$$

For each potential transition, we draw a duration. The shortest duration determines which transition occurs. For example, the duration until marriage for a never married person may be 5 years, while the duration until death may be 30 years. We conclude that the person will marry first. He now becomes subject to the competing hazards of divorce, widowhood, and death. Note that the mortality hazard is a function of marital status; now that the person has married, he faces more favorable survival chances. We therefore draw a new duration until death, taking account of the married status. In addition, durations are drawn until divorce and until the spouse's date of death. Whichever of the three randomly drawn durations comes first determines the next transition. This process continues until the person becomes deceased.[16]

2.7.1. Information from Numident Files

The last interview waves of the 1990 and 1991 SIPP panels took place sometime in 1992-94. Between then and June 1998, we know with certainty from administrative Numident records that some individuals deceased. The same issue arises with 1992 and 1993 SIPP respondents, whose Numident records are up-to-date through October 1998. Unfortunately, the Numident file is not complete, i.e., some individuals may have become deceased without corresponding record in the Numident file. Consider Figure 2.6, which graphs the natural logarithms of mortality rates based on Numident information (with 1990 and 1991 SIPP respondents as denominator) and 1994 Vital Statistics.

As is clear from visual inspection, death rates from Numident records are lower than they should be according to Vital Statistics. In other words, we can project deaths from Numident records with certainty, but we must generate additional deaths between the last survey wave and June 1998, when the Numident records were created. To this end, SAS program file `match.sas` generates mortality projections through June 1, 1998 (for 1990-91 SIPP respondents), and October 1, 1998 (1992-93 SIPP respondents). It finds that 2.6 percent of the sample (1,869 individuals) should be deceased as of 6/1/1998 or 10/1/1998, but that Numident records only show a death rate of 2.0 percent (1,444 individuals). In other words, the Numident records only appear to cover 77 percent (1,444/1,869) of the SIPP population.

[16] An alternative approach would project respondents' life paths in discrete steps, such as months. For example, the probability that a never married man marries during the next month is *p*; if this probability exceeds a randomly drawn variable (from a uniform distribution between 0 and 1), we project the wedding to occur. Similarly, project whether a death occurs and select the dominant transition. Then repeat for each subsequent month until death. This approach may be more suitable in models which incorporate covariates that vary frequently with time. It has limitations where multiple transitions are projected without a clearly dominant one, such as divorce and widowhood.

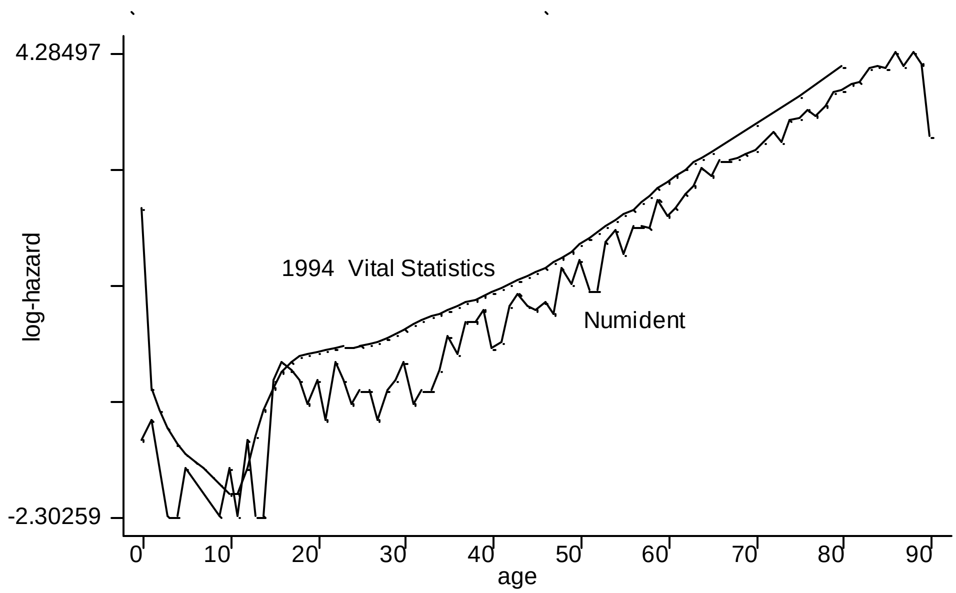

Figure 2.6. Log Death Rates, Numident vs 1994 Vital Statistics

We therefore randomly assign 77 percent of the SIPP sample as "matched" and 23 percent as non-matched. (This is random, except that the 1,444 deceased individuals are matched with certainty.) The projection method in `mint.sas` then distinguishes three types of individuals:

- 1,444 individuals are deceased with certainty in the month indicated by the Numident records.
- The remaining of the 77 percent "matched" respondents are guaranteed to survive through 6/1/1998 or 10/1/1998; after that date, the program accepts randomly generated survival durations.
- For the 23 percent non-matched respondents, the normal duration projections apply at all dates after the last interview wave.

It should be noted that the Numident records are subject to imperfect data quality. SSA staff matched the SIPP surveys to SSA's Master Numident file, and provided RAND with four small Numident files, for 1990, 1991, 1992, and 1993 SIPP respondents. These files contained a total of 10,228 records, each representing one deceased respondent. In many cases, the Numident month of death occurred before the final interview wave. Most of these indicated that the respondent died shortly before the last interview. We accepted the Numident information as correct provided that the Numident date was three months or less before the last interview. In 727 cases, the Numident date occurred more than three months before the last interview date; we assumed that an error was made in SSA's matching procedure and ignored Numident information for these respondents. In 39 cases, the Numident ID could not be matched to a SIPP individual. Again, we assumed that incorrect cases were pulled from the Master Numident file, and ignored Numident information for these 39 cases.

In six cases, there were duplicate Numident records. We randomly selected one Numident record from each pair and ignored the other information.

2.7.2. Characteristics of Future Spouses

Every time a respondent is projected to marry or remarry, a new spouse needs to be taken into consideration. Newly entering spouses do not become observations themselves; they only appear as new elements of variable arrays with spousal characteristics. The simulation database contains characteristics of every spouse, whether they are relevant before, during, or after the SIPP interviews. The characteristics include spousal sex, date of birth, race, ethnicity, education, disability status and date of disability onset. These characteristics are only known with certainty for spouses who were themselves respondents to the SIPP surveys. Even if they participated in only one interview, we recorded their characteristics. Above we explained how we imputed characteristics of former spouses. Characteristics of future spouses are imputed in exactly the same manner. They are directly relevant for the projection of spousal death dates (widowhood dates).

2.7.3. Spousal Consistency

Our projection algorithms are designed to ensure spousal consistency: if a couple is married as of the last interview date, we project the next transition to be on the same date for husband and wife. If the first transition is a divorce, we project the same divorce date for husband and wife; if the first transition is the husband's death, we project that the wife becomes widowed on that same date; and similarly, we project his widowhood date to be at her death date.

Spousal consistency is achieved by using the same random number seed for husbands and wives. Three potential transitions are relevant: divorce, his death, and her death. His death only involves his characteristics, i.e., respondent characteristics when processing his projections, and spousal characteristics when projecting her future. The mirror case arises for her date of death, i.e., his widowhood date. Projections of the divorce date generate an additional complexity because divorce equations are estimated separately for males and females (Table 2.7, page 17).[17] If we were to use respondents' own characteristics, different divorce dates would be generated, even if the seeds were equal. In light of indications that women's marriage history reports tend to be of higher quality than men's, we project divorce dates based on the wife's characteristics (and the female divorce model coefficients), if available. In other words, projections of a divorce date of a woman are always based on her own characteristics and the female divorce model. Projections of a divorce date of a man

[17] It is impossible to estimate divorce equations using both the husband's and the wife's characteristics, including measures of spousal compatibility such as whether they are of the same race, because spousal characteristics are only known for marriages that are still ongoing at the time of the SIPP interviews. Characteristics of former spouses are imputed and thus contain only noise.

are based on his wife’ s characteristics if these are known with certainty, i.e., for marriages that were in progress at the last interview date. Divorce dates of men’s future marriages are based on his own characteristics (and the male divorce specification).

2.8. Projection Results

The previous section explained how we project individuals' life course, starting at the last interview date and ending at the date of death. We generate variables and variable arrays to record all transitions: array `marb` for marriage begin dates; array `mare` for marriage end dates; array `howend` for the type of marriage disposition (divorce, widowhood, own death); variable `disabled` for whether the person became disabled before death; variable `disabdte` for the onset of disability (if `disabled`=1); and variable `deathdte` for the date of death.

SAS macro `%figstat` may be used to determine individual' s demographic status at any particular date. The following tables show the projected demographic distribution as of January 1, 2020. Table 2.10 tabulates demographic status for all 1990-93 SIPP respondents born in 1931-60; Table 2.11 is conditional on survival through 2020.

Table 2.10. Projected Demographic Distribution in 2020 (in percent; 1931-60 birth cohort)

	Male	Female	Total
Never married	3.5	5.0	4.2
Married	47.3	39.7	43.4
Widowed	3.7	19.6	11.8
Divorced	8.2	14.9	11.6
Deceased	37.4	20.9	29.0
Total	100.0	100.0	100.0

Table 2.11. Projected Demographic Distribution in 2020 (in percent; 1931-60 birth cohort, survivors only)

	Male	Female	Total
Never married	5.5	6.3	6.0
Married	75.6	50.2	61.2
Widowed	5.8	24.8	16.6
Divorced	13.1	18.8	16.3
Total	100.0	100.0	100.0

We project that 29 percent of the respondents in the simulation sample will be deceased by the year 2020. Almost twice as many men as women will have become deceased. Among the survivors, 6 percent will have never married, 61 percent will be married, whereas the remaining one-third will be equally divided between divorced and widowed. However, there will be many fewer widowers than widows. Only about 6 percent of surviving men will be widowed, compared with one out of four surviving women.

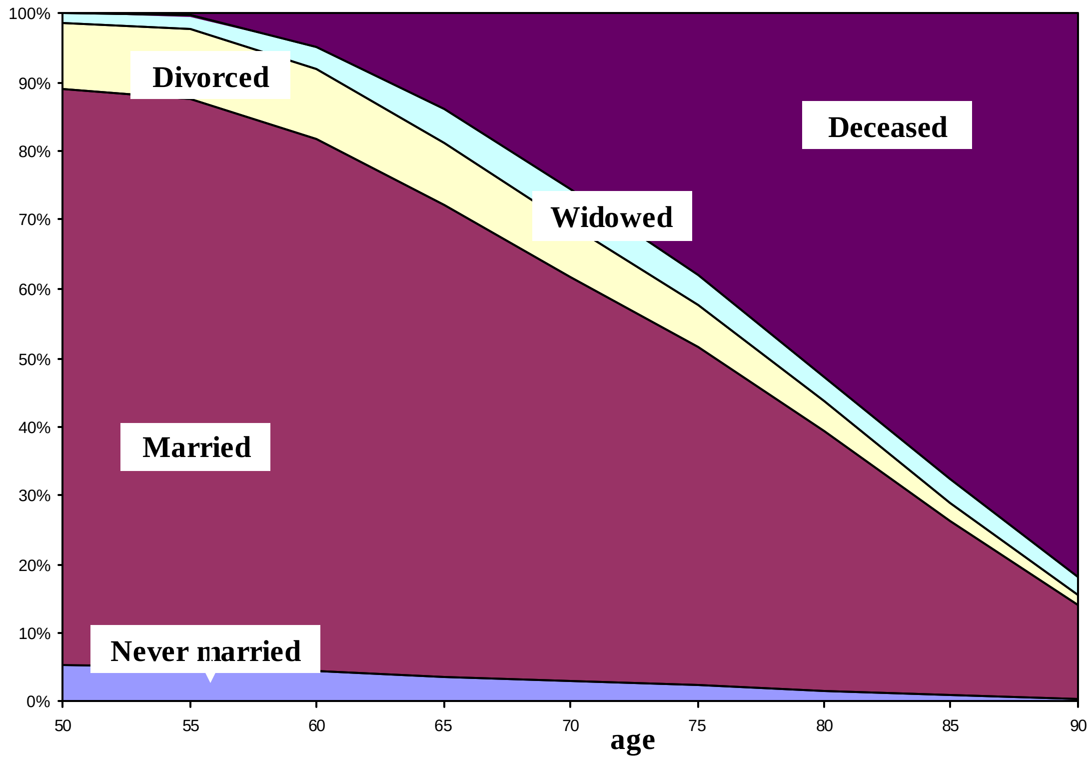

Figure 2.7. Life Cycle Composition: Men Born in 1931-40

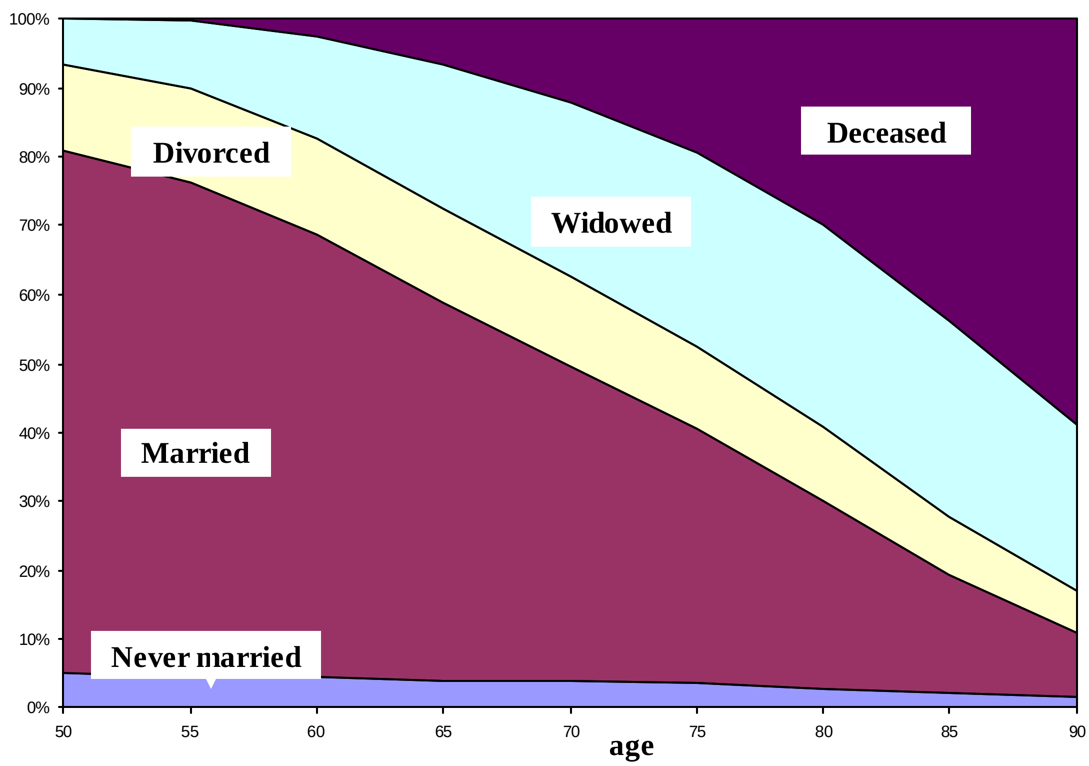

Figure 2.8. Life Cycle Composition: Women Born in 1931-40

Figure 2.7 and Figure 2.8 show the life cycle composition of men and women born in 1931-40, respectively. The figures start at age 50, when the youngest in this cohort participate in SIPP interviews. As of age 50, all are thus alive to participate in SIPP surveys. The upper bound is age 90, corresponding to the year 2021 for the oldest individuals. As individuals age, an increasing number becomes deceased or widowed. Note the large differences between men and women: men remain overwhelmingly married, but as their numbers become smaller, a large fraction of women becomes widowed.

Figure 2.9 and Figure 2.10 show life cycle demographic compositions of surviving men and women born in 1931-1940, i.e., similar to the previous two figures but without the deceased category. The distribution of men by marital status remains virtually unchanged, with predominantly (re)married men. Women, on the other hand, become increasingly widowed at advanced ages.

An important reason for developing the MINT microsimulation model, as opposed to a macro model, is the ability to determine program eligibility for individuals based on individuals' unique characteristics. Consider Figure 2.10, which shows that about 14.5 percent of women born in 1931-40 reach age 62 as divorcées. What fraction of these women will be able to claim Social Security benefits on the basis of their ex-husbands' earnings? Of all individuals that reach age 62 as in divorced status, Table 2.12 shows the fraction whose most recent marriage lasted less than ten years. Overall, about 39 percent of divorced women reach age 62 without a claim on spousal benefits. (In addition, 43 percent of divorced men cannot claim spousal benefits, but they are more likely to have had substantial earnings themselves.) The ineligible fraction is increasing by birth cohort. *Overall, 3.2 million divorced women in the 1931-60 cohort will not be eligible for spousal benefits.* To determine how many of these women would have had sufficiently low lifetime earnings so as to collect spousal benefits, one needs to consider the earnings projections as produced by The Urban Institute/Brookings Institution.

Table 2.12. Fraction Divorced Individuals Married Less Than Ten Years

	Male	Female	Total
1931-40 cohort	40.9	31.1	34.9
1941-50 cohort	42.9	39.9	41.1
1951-60 cohort	44.1	41.4	42.5
Total	43.1	38.9	40.6

A similar calculation may be carried out to determine the fraction of widows ineligible for widowhood benefits because they were married less than nine months (Social Security Handbook §401). MINT projects that 30.2 million women in the 1931-60 birth cohorts become widowed. Of these, about 220,000 (0.7 percent) were married less than nine months. In addition, about 40,000 men became widowed less than nine months after their wedding.

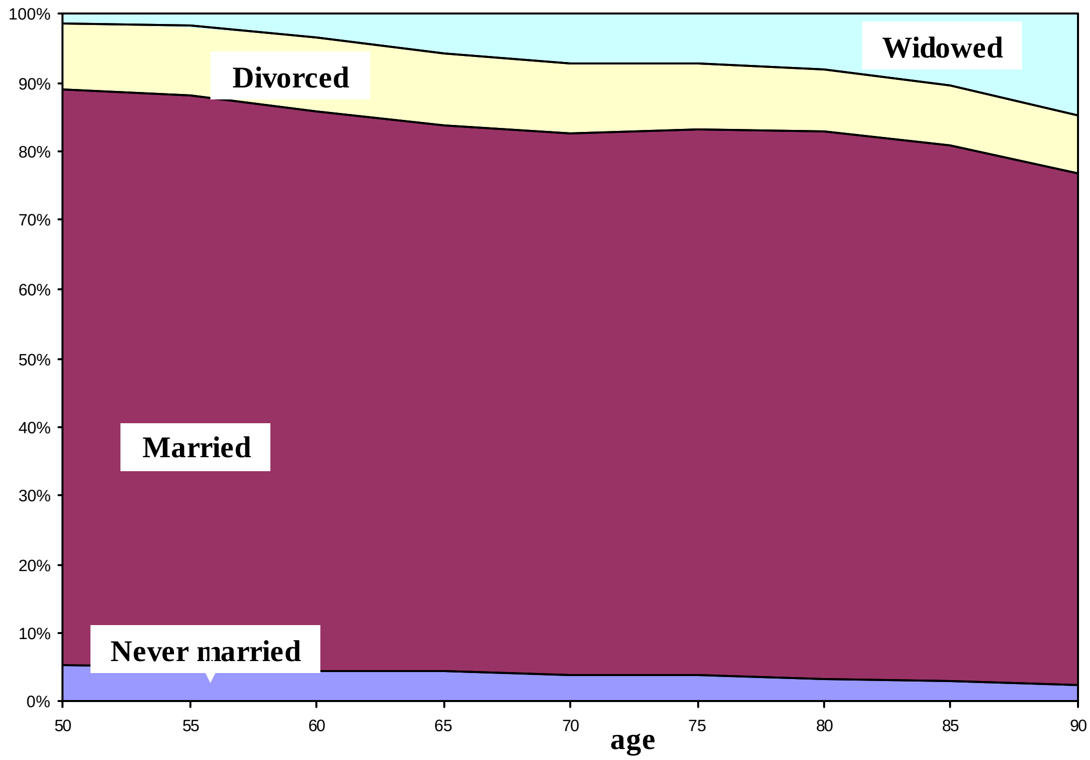

Figure 2.9. Life Cycle Composition: Surviving Men Born in 1931-40

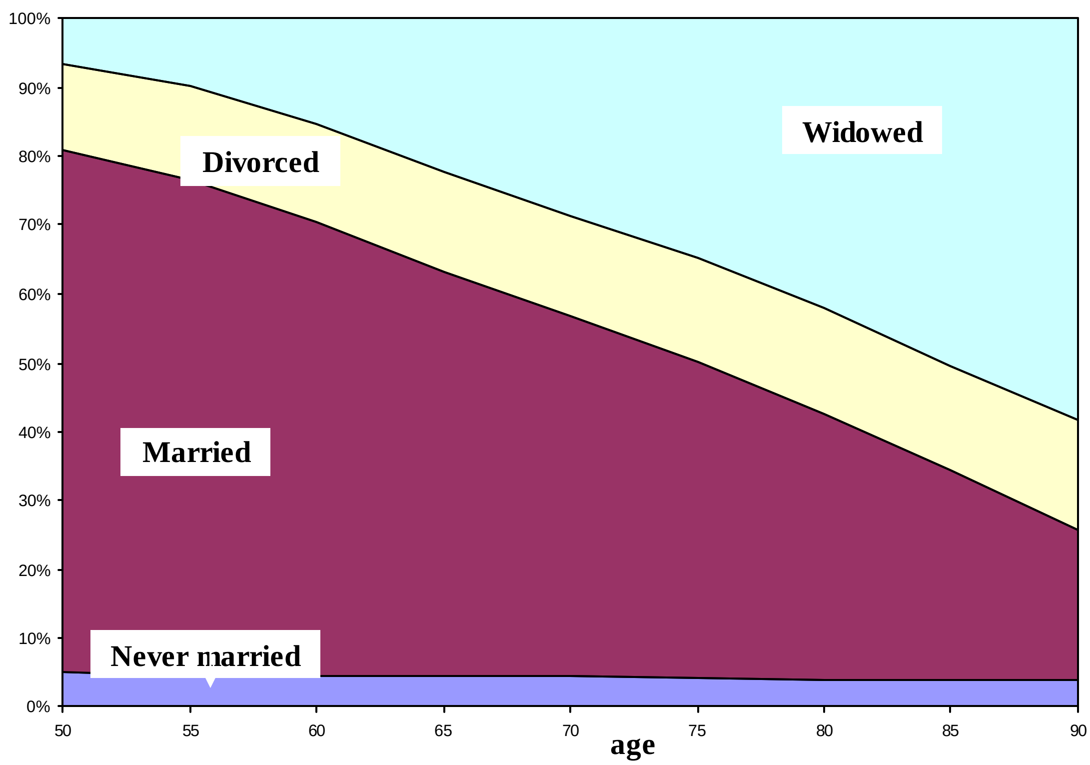

Figure 2.10. Life Cycle Composition: Surviving Women Born in 1931-40

Chapter 4 returns to our demographic projections and compares them to those produced by SSA's Office of the Chief Actuary.

2.9. Mortality as a Function of Disability Status

As documented above, survival projections are based on a mortality hazard model which does not account for disability status. However, disability status is a strong predictor of survival, as shown in Table 2.13. Disabled males face mortality risks that are 2.45 (=exp(0.8971)) times as high as those experienced by their disability-free counterparts, whereas disability increased women' s mortality risk by a factor of 2.94 (=exp(1.0816)).[18]

Note that the effect of permanent income on mortality risk is substantially smaller than in the specification without control for disability status (Table 2.1).

The projection algorithms support longevity projections which take account of disability status. See Appendix A.7 for an explanation of how to modify the projection program such that the projections are based on the specification with account of disability status of Table 2.13.[19]

[18] The PSID did not collect disability status of wives in earlier waves. Married women that became deceased early in the panel thus often have missing disability status, which explains the positive and significant coefficient on missing disability status for women.

[19] It should be noted that account for disability status requires more than modification of the longevity projection algorithms. In particular, income from assets as projected by The Urban Institute assumes that families purchase an lifelong joint and survivor annuity with 80 percent of their assets. The current annuitization algorithms do not take account of disability status of husbands or wives. They need to be modified for consistency throughout all MINT components, such that the disability-free face less generous annuity tables than the disabled.

Table 2.13. PSID Mortality Hazard Estimates, Controlling for Disability Status

	Males	Females
Constant	-9.4700 ***	-10.0894 ***
	(0.2626)	(0.3428)
Age slope 30-65	0.0758 ***	0.0774 ***
	(0.0046)	(0.0058)
Age slope 65+	0.0679 ***	0.0787 ***
	(0.0043)	(0.0048)
Calendar time	-0.0156 ***	-0.0229 ***
	(0.0039)	(0.0051)
Black	0.1810 **	0.2560 ***
	(0.0811)	(0.0962)
High school drop-out	0.3350 ***	0.0515
	(0.0710)	(0.0778)
College graduate	0.0139	-0.2422 *
	(0.1032)	(0.1404)
Never married	0.2725 **	-0.0609
	(0.1130)	(0.1442)
Divorced	0.3525 ***	-0.2248
	(0.1134)	(0.1564)
Widowed	0.1305	-0.1530 *
	(0.0922)	(0.0843)
Disabled	0.8971 ***	1.0816 ***
	(0.0779)	(0.1032)
Disability status missing	0.0815	0.5158 ***
	(0.3591)	(0.1387)
Permanent income	-0.0851 *	-0.1920 ***
	(0.0441)	(0.0499)
Income missing	-0.4138	-1.9734
	(1.1782)	(3.9264)
Log-Likelihood	-14287.14	

Note: asymptotic standard errors in parentheses;
significance `*' = 10%, `**' = 5%, `***' = 1%